

# A computational method for bounding the probability of reconstruction on trees

Nayantara Bhatnagar\*

Elitza Maneva†

March 27, 2009

## Abstract

For a tree Markov random field non-reconstruction is said to hold if as the depth of the tree goes to infinity the information that a typical configuration at the leaves gives about the value at the root goes to zero. The distribution of the measure at the root conditioned on a typical boundary can be computed using a distributional recurrence. However the exact computation is not feasible because the support of the distribution grows exponentially with the depth.

In this work, we introduce a notion of a *survey* of a distribution over probability vectors which is a succinct representation of the true distribution. We show that a survey of the distribution of the measure at the root can be constructed by an efficient recursive algorithm. The key properties of surveys are that the size does not grow with the depth, they can be constructed recursively, and they still provide a good bound for the distance between the true conditional distribution and the unconditional distribution at the root. This approach applies to a large class of Markov random field models including randomly generated ones. As an application we show bounds on the reconstruction threshold for the Potts model on small-degree trees.

---

\*Department of Statistics, University of California at Berkeley. (nayan@stat.berkeley.edu) Supported by DOD ONR grant N0014-07-1-05-06 and DMS 0528488.

†Universitat Politècnica de Catalunya (maneva@lsi.upc.edu); work done in part while this author was a postdoctoral researcher at IBM Almaden.

# 1 Introduction

Correlations between distant elements or sets of elements in randomly generated Markov random fields (MRFs) are a main consideration in the analysis of random constraint satisfaction problems, and in the design and analysis of message-passing, local search, and other iterative algorithms for these problems. Here we study one such concept of correlation on tree MRFs. The presence of this correlation is known as the property of *reconstruction*, and its absence as *non-reconstruction*. Non-reconstruction is equivalent to the free-boundary Gibbs measure of the tree being *extremal* [5]. From the point of view of statistical physics, reconstruction is equivalent to *replica-symmetry breaking* [9]. Some important results on reconstruction are [6, 3, 11, 12, 2, 1, 15, 16]. For the connection between this concept and the design of algorithms for constraint satisfaction and optimization problems we refer the reader to [10, 17, 8] and the references therein.

Our contribution is the first general efficient computational method to obtain non-trivial bounds for the probability of reconstruction for a tree Markov random field. We illustrate the method with an application to the Potts model.

Consider a tree MRF. The distribution of interest is the conditional distribution at the root, given a boundary that is generated randomly according to the MRF. It is said that the root *cannot be reconstructed*, or the non-reconstruction property holds, if and only if with high probability this distribution converges to the unconditional marginal distribution at the root as the depth of the tree goes to infinity. The distribution of the conditional distribution at the root (conditioned on a random boundary) can be expressed recursively<sup>1</sup> - there is a simple analytic expression for the distribution for a tree of depth  $n + 1$  in terms of the distribution for the tree of depth  $n$ . However as the depth of the tree increases the support of the distribution grows exponentially, which makes numerical estimates difficult to obtain. For analytical analysis one essentially needs to find a special parameter of the distribution which can be bounded recursively, and this depends on the particular MRF. Often the analysis has two steps - first showing that the expected distance from the unconditional distribution is below some small constant, and then showing that arriving below this small constant implies that this distance (or some parameter related to it) decreases geometrically. Particular examples of this approach are the recent results on reconstruction for colorings of Bhatnagar, Vera, Vigoda, and Weitz [1] and Sly [15], and for the q-state Potts model of Sly [16]. The moral from their analysis is that the first step is the more difficult to achieve. In particular, it is the step that makes the analysis possible only for the case of large-degree trees. Here we propose a very general method for making this first step, independent of the particular method for generating the MRF, and which is practical for the case of small-degree trees.

We introduce the notion of a *survey* of a distribution over probability vectors. The survey can be thought of as a "projection" onto a small "basis" set of probability vectors that carries enough information about the true distribution. In particular, for the reconstruction problem, when the goal is to bound from above the probability of reconstruction, we show that it suffices to keep a survey of the distribution at each iteration. Applying the recursion to the surveys for enough iterations allows us to obtain very good (possibly arbitrarily good) bounds on the probability of reconstruction.

We apply our method to the symmetric 3-state Potts model with various parameters in order to compare with previous results, although it will be clear that the method does not depend on the symmetry of the model.

---

<sup>1</sup>Notice that the "distribution of ... distribution" is not a mistake - the object we are interested in is indeed the distribution of the randomness left in the root after looking at the boundary.

For the Potts model the second step is also easy to achieve using the ideas from recent work of Sly [16]. Since the complexity of the method is exponential in the degree of the tree, and in the alphabet of the MRF, we do not apply it to very large parameters. However, we are able to demonstrate new bounds for 3-spins on the 2-ary and 3-ary tree, and on a random tree in which every internal vertex has either 2 or 3 children with equal probability, improving the bounds of Mossel and Peres [12]. We should point out that very recently Formentin and Külske [4] demonstrated even better results for this model. Thus, rather than the numerical results, the importance of our contribution is in the generality of our method.

The algorithm we present can be viewed as a rigorous alternative to the population dynamics algorithm used in physics [9, 17] to determine the spin-glass transition. In population dynamics the distributional recursion is approximated by keeping a sample of the distribution at each step. What is not known rigorously is whether applying the recursion to a sample of the distribution for the tree of depth  $n$  really results in some sense in a “faithful” sample of the distribution for the tree of depth  $n + 1$ . In contrast, the main technical lemma of our work is precisely the statement that applying the recursion to the survey of the distribution for the tree of depth  $n$  results in a survey of the distribution for the tree of depth  $n + 1$ .

Another related algorithm is the density evolution algorithm for analysis of the probability of bit-error of LDPC codes [13, 14]. The density evolution algorithm, as its name indicates, is also a recursion on distributions, and in practice it is carried out by heuristically quantizing the distribution at every step by rounding. Unfortunately, unlike the reconstruction recursion, the density evolution recursion does not commute with taking surveys of the distributions therefore our method cannot be applied, or at least not in the obvious way.

Quantization of distributions is also an important design step in the Survey Propagation algorithm of [10]. In its analytical form Survey Propagation uses messages that are distributions with growing support, whereas in its practical form the messages are distributions with support of size 3. However we do not know if there is a precise mathematical connection between the two kinds of quantization.

This article is organized as follows. In the next section we give the technical definitions related to reconstruction, and the recursion relation that we analyze. Section 3 is dedicated to the new definition of a survey of a distributions, its key properties, and how it can be applied to the reconstruction problem. In the last section we discuss an application to the Potts model, and compare the results obtained by our method to previous bounds.

## 2 Reconstruction on trees

The reconstruction problem in its simplest form can be stated in terms of a broadcast problem. Consider a process in which information is broadcast from the root of an infinite rooted tree  $T$  to other vertices as follows. Each edge  $e = (u, v)$  acts as a channel  $M$  with a finite alphabet  $D = \{1, 2, \dots, q\}$ . The channel  $M$  is a Markov chain where  $(M)_{i,j} = \Pr(v = j | u = i)$ .

The letter at the root  $\rho$ , denoted by  $\sigma_\rho$ , is chosen according to some initial distribution. This value is then propagated in the tree as follows. For vertex  $v$  with parent  $u$ , let  $\sigma_v = M(\sigma_u)$  for each edge independently.

For distributions  $\mu$  and  $\nu$  on the same space  $\Omega$ , the *total variation distance* between  $\mu$  and  $\nu$  is

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{\sigma \in \Omega} |\mu(\sigma) - \nu(\sigma)|$$

Let  $L(n)$  denote the configuration at level  $n$ .

**Definition 1** *The tree  $T$  and channel  $M$  have the non-reconstruction property if for every  $a, b \in D$ ,*

$$\lim_{n \rightarrow \infty} d_{TV}[L^a(n), L^b(n)] = 0$$

where  $L^c(n)$  denotes the conditional distribution of  $L(n)$  given that  $\sigma_\rho = c$ .

In 2006 Mézard and Montanari [9] showed that the same problem had also been studied in physics for even more general models, and the reconstruction transition is the equivalent to the *replica symmetry breaking spin-glass transition*. In the physics formalism rather than using channels, the system is described as a Markov random field, which in many application is also randomly generated, and the question again is whether the root is reconstructible from typical assignments at the leafs.

Here, for greatest generality we will define the reconstruction problem using a Markov random field (MRF). This way we capture models such as 3-SAT and other constraint satisfaction problems, which are not as easy to think about in terms of channels.

Consider a tree  $T = (V, E)$  with root vertex  $\rho$ . For every  $v \in V$  there is a domain of values  $D_v$  that can be assigned to this vertex, and for every edge  $e = (u, v)$  a non-negative function we call potential  $\Psi_e : D_u \times D_v \rightarrow \mathbb{R}^+$ . We will define a distribution over assignments to the vertices  $\times_{v \in V} D_v = D$ . For every configuration  $\bar{\sigma} \in D$ , let  $\sigma_v$  denote the component corresponding to vertex  $v$ , and  $\bar{\sigma}(n)$  the components corresponding to level  $n$  of the tree.

$$\mathbb{P}_{T, \Psi}(\bar{\sigma}) := \frac{1}{Z_{T, \Psi}} \prod_{e=(u,v) \in E} \Psi_e(\sigma_u, \sigma_v), \quad (1)$$

where  $Z$  is the normalizing constant  $Z_{T, \Psi} = \sum_{\bar{\sigma} \in D} \prod_{e=(u,v) \in E} \Psi_e(\sigma_u, \sigma_v)$ .

We will allow random MRFs  $(T, \Psi)$  generated in the following way. To each level there corresponds a degree distribution and a distribution over potential functions. A tree of depth 1 is just a single vertex (no potential functions). An MRF of depth  $n + 1$  is generated by choosing a degree for the root from the degree distribution of level  $n + 1$ , next choosing potential functions for all the edges adjacent to the root independently from the distribution of potential functions corresponding to level  $n + 1$ , and finally attaching randomly generated MRFs of depth  $n$  to the other ends of the edges.

We denote by  $L(n)$  the configuration at the vertices at level  $n$  (assuming  $T$  is of depth at least  $n$ ), and by  $\rho$  the root. Let also  $L^a(n)$  denote the configuration at level  $n$  conditional on the root having value  $a \in D_\rho$ .

**Definition 2** *We say that a random MRF has the non-reconstruction property if for every  $a, b \in D_\rho$*

$$\lim_{n \rightarrow \infty} \mathbb{E}_{T, \Psi}[d_{TV}(L^a(n), L^b(n))] = 0.$$

For the rest of this section we will consider a fixed MRF, so we can omit  $\Psi$  from the subscript as the functions on the edges will not be changing, and we will also omit  $T$  from the subscript whenever it is understood. We will also use  $a$  and  $b$  for the events that the root takes the value  $a$  or  $b$  respectively.

We will use an alternative expression for the total variation distance, which follows from Bayes' rule. For this we denote  $\pi(a) := \mathbb{P}_T(a)$  and  $\pi(b) := \mathbb{P}_T(b)$ .

$$d_{TV}(L^a(n), L^b(n)) = \sum_L |\mathbb{P}(L(n) = L|a) - \mathbb{P}(L(n) = L|b)| \quad (2)$$

$$= \sum_L \left| \frac{\mathbb{P}(a|L(n) = L)\mathbb{P}(L(n) = L)}{\pi(a)} - \frac{\mathbb{P}(b|L(n) = L)\mathbb{P}(L(n) = L)}{\pi(b)} \right| \quad (3)$$

$$= \mathbb{E}_{L(n)} \left[ \left| \frac{\mathbb{P}(a|L(n))}{\pi(a)} - \frac{\mathbb{P}(b|L(n))}{\pi(b)} \right| \right] \quad (4)$$

$$\leq \frac{1}{\pi(a)} \mathbb{E}_{L(n)} [|\mathbb{P}(a|L(n)) - \pi(a)|] + \frac{1}{\pi(b)} \mathbb{E}_{L(n)} [|\mathbb{P}(b|L(n)) - \pi(b)|]. \quad (5)$$

An immediate observation is that this quantity is monotonically decreasing with the depth of the tree  $n$ .

$$\begin{aligned} & \mathbb{E}_{L(n+1)} \left[ \left| \frac{\mathbb{P}(a|L(n+1))}{\pi(a)} - \frac{\mathbb{P}(b|L(n+1))}{\pi(b)} \right| \right] \\ &= \mathbb{E}_{L(n+1)} \left[ \left| \sum_L \frac{\mathbb{P}(a|L(n) = L) \mathbb{P}(L(n) = L|L(n+1))}{\pi(a)} - \frac{\mathbb{P}(b|L(n) = L) \mathbb{P}(L(n) = L|L(n+1))}{\pi(b)} \right| \right] \\ &\leq \mathbb{E}_{L(n+1)} \left[ \sum_L \mathbb{P}(L(n) = L|L(n+1)) \times \left| \frac{\mathbb{P}(a|L(n) = L)}{\pi(a)} - \frac{\mathbb{P}(b|L(n) = L)}{\pi(b)} \right| \right] \\ &= \sum_{L'} \mathbb{P}(L(n+1) = L') \sum_L \mathbb{P}(L(n) = L|L(n+1) = L') \times \left| \frac{\mathbb{P}(a|L(n) = L)}{\pi(a)} - \frac{\mathbb{P}(b|L(n) = L)}{\pi(b)} \right| \\ &= \sum_L \mathbb{P}(L(n) = L) \left| \frac{\mathbb{P}(a|L(n) = L)}{\pi(a)} - \frac{\mathbb{P}(b|L(n) = L)}{\pi(b)} \right| \\ &= \mathbb{E}_{L(n)} \left[ \left| \frac{\mathbb{P}(a|L(n))}{\pi(a)} - \frac{\mathbb{P}(b|L(n))}{\pi(b)} \right| \right]. \end{aligned}$$

This expression for the variation distance is a function of a distribution that we will refer to many times in the rest of the paper, so for convenience we define the following terminology for it:

**Definition 3** For a tree Markov random model as in equation (1) the residual distribution at the root of  $T$  is the distribution of the marginal distribution at the root conditional on a random boundary, which is chosen according to the distribution  $\mathbb{P}_{T,\Psi}$ . In other words it is the distribution of the probability vector  $\eta = (\eta^a := \mathbb{P}_T(a|L(n)), a \in D_\rho)$ , where  $L(n)$  is a configuration for the leaves chosen according to  $\mathbb{P}_{T,\Psi}$ .

Intuitively, a random MRF has the non-reconstruction property if the residual distribution at the root is with high probability concentrated on probability vectors arbitrarily close to the stationary distribution  $\pi$ . In particular, we will aim to show that  $\mathbb{E}[|\mathbb{P}(a|L(n)) - \pi(a)|]$  goes to 0 for every  $a \in D_\rho$ , which implies non-reconstruction by the inequality (5).

Next we derive the recursive equation for the residual distribution at the root of a tree. For a tree  $T$  of depth  $n$  with edges  $E_T$  let

$$Z_T(\vec{\sigma}) := \prod_{e=(u,v) \in E_T} \Psi_e(\sigma_u, \sigma_v).$$

Thus  $Z_T := \sum_{\bar{\sigma}} Z_T(\bar{\sigma})$ . For a boundary configuration  $L$ , let

$$Z_T(L) := \sum_{\bar{\sigma}: \bar{\sigma}(n)=L} Z_T(\bar{\sigma}).$$

For a boundary  $L$  let  $\eta_T(L) = (\eta_T^a(L) : a \in D_\rho)$  denote the probability vector for the distribution of the assignment at the root conditional on the boundary being  $L$ , i.e.  $\eta_T^a(L) := \mathbb{P}_T(a|L(n) = L)$ . For any probability vector  $\eta$  on  $D_\rho$  let

$$Z_T(\eta) := \sum_{L: \eta_T(L)=\eta} Z_T(L).$$

Thus the probability that the marginal distribution at the root is equal to  $\eta$  is  $Z_T(\eta)/Z_T$ .

Suppose the children of the root of  $T$  are  $v_1, \dots, v_r$ , the corresponding edges connecting them to the root are  $e_1, \dots, e_r$  and the subtrees rooted at them are respectively  $T_1, \dots, T_r$ . Consider a boundary configuration  $L = (L_1, \dots, L_r)$  for the large tree  $T$ , where  $L_i$  is the part of the boundary that belongs to  $T_i$ . It is straightforward to derive the expression for  $\eta_T(L)$  in terms of  $\eta_{T_i}(L_i)$ :

$$\begin{aligned} \eta_T^a(L) &= \frac{\sum_{\bar{\sigma}: \bar{\sigma}(n)=L, \sigma_\rho=a} Z(\bar{\sigma})}{Z_T(L)} \\ &= \frac{\sum_{\sigma_{v_1}, \dots, \sigma_{v_r}} \prod_{i=1}^r \Psi_{e_i}(a, \sigma_{v_i}) Z_{T_i}(L_i) \eta_{T_i}^{\sigma_{v_i}}(L_i)}{Z_T(L)} \\ &= \frac{\prod_{i=1}^r Z_{T_i}(L_i)}{Z_T(L)} \prod_{i=1}^r \sum_{\sigma_{v_i} \in D_{v_i}} \Psi_{e_i}(a, \sigma_{v_i}) \eta_{T_i}^{\sigma_{v_i}}(L_i) \end{aligned}$$

It is convenient to define the corresponding function (which is actually the update function of the belief propagation algorithm for computing the marginal distribution at the root of an acyclic MRF). Let  $\mathcal{P}_v$  be the space of probabilities over the domain  $D_v$  for any  $v \in V$ . We define  $f : \mathcal{P}_{v_1}, \dots, \mathcal{P}_{v_r} \rightarrow \mathcal{P}_\rho$  in the following way:

$$f^a(\eta_1, \eta_2, \dots, \eta_r) := \prod_{i=1}^r \sum_{\sigma_{v_i} \in D_{v_i}} \eta_i^{\sigma_{v_i}} \Psi_{e_i}(a, \sigma_{v_i}).$$

We will also need to use the norm  $\|f(\eta_1, \dots, \eta_r)\| = \sum_{a \in D_\rho} f^a(\eta_1, \dots, \eta_r)$ . With this notation

$$\eta_T^a(L) = \frac{f^a(\eta_{T_1}(L_1), \dots, \eta_{T_r}(L_r))}{\|f(\eta_{T_1}(L_1), \dots, \eta_{T_r}(L_r))\|}.$$

For two vectors the symmetric relation  $\propto$  will indicate that the vectors are equal up to a multiplicative constant. Thus  $\eta_T(L) \propto f(\eta_{T_1}(L_1), \dots, \eta_{T_r}(L_r))$ . When used with probabilities it will indicate that the normalization constant has been omitted. We are now ready to derive the recursion for the residual distribution at the root of the tree.

**Theorem 4** *Let  $\mathbf{P}_i$  and  $\mathbf{Q}$  be random vectors such that  $\mathbf{P}_i$  is distributed according to the residual distribution at the root of  $T_i$  for  $i = 1, \dots, r$  and  $\mathbf{Q}$  is distributed according to the residual distribution at the root of  $T$ . Then for any  $\eta \in \mathcal{P}_\rho$*

$$\mathbb{P}(\mathbf{Q} = \eta) \propto \mathbb{E}[\|f(\mathbf{P}_1, \dots, \mathbf{P}_r)\| \times \mathbb{I}nd[f(\mathbf{P}_1, \dots, \mathbf{P}_r) \propto \eta]].$$

**Proof:** First we derive the recursion for the total weight of configurations with a fixed boundary:

$$\begin{aligned}
Z_T(L) &= \sum_{a \in D_\rho} \sum_{\sigma_{v_1}, \dots, \sigma_{v_r}} \prod_{i=1}^r \Psi_{e_i}(a, \sigma_{v_i}) Z_{T_i}(L_i) \eta_{T_i}^{\sigma_{v_i}}(L_i) \\
&= \left( \prod_{i=1}^r Z_{T_i}(L_i) \right) \times \sum_{a \in D_\rho} \prod_{i=1}^r \sum_{\sigma_{v_i} \in D_{v_i}} \eta_{T_i}^{\sigma_{v_i}}(L_i) \Psi_{e_i}(\sigma_\rho, \sigma_{v_i}) \\
&= \left( \prod_{i=1}^r Z_{T_i}(L_i) \right) \times \|f(\eta_{T_1}(L_1), \dots, \eta_{T_r}(L_r))\|.
\end{aligned}$$

Next, using the above, we derive the recursion for the total weight of configurations yielding a given marginal distribution at the root:

$$\begin{aligned}
Z_T(\eta) &= \sum_{L: \eta_T(L) = \eta} Z_T(L) \\
&= \sum_{\substack{L = (L_1, \dots, L_r): \\ f(\eta_{T_1}(L_1), \dots, \eta_{T_r}(L_r)) \propto \eta}} \left( \prod_{i=1}^r Z_{T_i}(L_i) \right) \times \|f(\eta_{T_1}(L_1), \dots, \eta_{T_r}(L_r))\| \\
&= \sum_{\substack{\eta_1, \dots, \eta_r: \\ f(\eta_1, \dots, \eta_r) \propto \eta}} \|f(\eta_1, \dots, \eta_r)\| \sum_{L_1: \eta_{T_1}(L_1) = \eta_1} \dots \sum_{L_r: \eta_{T_r}(L_r) = \eta_r} \prod_{i=1}^r Z_{T_i}(L_i) \\
&= \sum_{\substack{\eta_1, \dots, \eta_r: \\ f(\eta_1, \dots, \eta_r) \propto \eta}} \|f(\eta_1, \dots, \eta_r)\| \prod_{i=1}^r Z_{T_i}(\eta_i).
\end{aligned}$$

Finally we can derive the recursion for the residual distribution at the root:

$$\begin{aligned}
\mathbb{P}(\mathbf{Q} = \eta) &= \frac{Z_T(\eta)}{Z_T} \\
&= \frac{\prod_{i=1}^r Z_{T_i}}{Z_T} \sum_{\substack{\eta_1, \dots, \eta_r: \\ f(\eta_1, \dots, \eta_r) \propto \eta}} \|f(\eta_1, \dots, \eta_r)\| \prod_{i=1}^r \frac{Z_{T_i}(\eta_i)}{Z_{T_i}} \\
&= \frac{\prod_{i=1}^r Z_{T_i}}{Z_T} \mathbb{E}[\|f(\mathbf{P}_1, \dots, \mathbf{P}_r)\| \times \text{Ind}[f(\mathbf{P}_1, \dots, \mathbf{P}_r) \propto \eta]].
\end{aligned}$$

■

This recursion was also derived by Mézard and Montanari in [9] and, as pointed out by them, its fixed point is known as the “1-step Replica Symmetric Breaking solution with Parisi parameter  $m = 1$ ” (in the general 1RSB scheme the factor  $\|f(\dots)\|$  is raised to a power  $m$ ). Almost all other work in reconstruction considers, instead, the recursion for the *conditional* residual distribution at the root, conditioned on the boundary being generated from the MRF with a fixed value at the root. The method we present here can be applied only with the unconditional distribution.

The main contribution of this article is a method for discretizing the recursion of Theorem 4. The only property of  $f$  that is used is that it is a multi-affine function (i.e. affine in each coordinate). Thus the main technical theorem will not use the definitions related to reconstruction.



### 3 Surveys of distributions

Let us first illustrate our approach with the example of the channel corresponding to random coloring on the regular  $d$ -ary tree. Each vertex can take one of  $q$  colors. Based on the color of the parent, the colors of the children are chosen independently and uniformly at random from the set of colors different from that of the parent. If we start with a uniformly random color at the root, this process generates a random coloring of the tree. Showing non-reconstruction for this process is equivalent to showing that for almost all colorings of the leafs generated by choosing a random coloring of the tree, there are almost the same number of colorings consistent with these leafs, in which the root of the tree has each of the  $q$  colors. More precisely, let  $C_L$  denote the colorings of a tree of depth  $n$  with leafs colored according to  $L$ , and  $C_L^i$  denote the colorings in  $C_L$  with color  $i$  at the root. We will recursively try to match up the colorings in the sets  $C_L^1, \dots, C_L^q$  by splitting them into sets that are balanced or close to balanced. For example, suppose  $q = 3$ , the colors are called Red, Blue and Green, and for some  $L$  we have  $(C_L^R, C_L^B, C_L^G) = (10, 5, 7)$ . Then we can split this set of colorings into 3 sets : one that is perfectly balanced, containing 5 colorings of each type; one that is balanced with respect to Red and Green, containing two colorings with Red at the root, and two with Green at the root; and one that contains the remaining 3 colorings with Red at the root  $((10, 5, 7) = (5, 5, 5) + (2, 0, 2) + (3, 0, 0))$ . We call such sets *bundles*. A bundle always contains only colorings that have the same colors at the leafs. Bundles can be of several different types according to the ratio of number of colorings with red, blue and green at the root, and we can choose the types that will be allowed. We will only keep track of the number of bundles of each type. The goal is to construct the bundles recursively in such a way that the majority of colorings are eventually in balanced bundles.

There is a simple process to construct bundles recursively. Suppose we have a particular splitting of the colorings of the  $d$ -ary tree of depth  $n$  into bundles. For any  $d$ -ple of these bundles, consider the set of colorings of the tree of depth  $n + 1$  such that the first subtree of depth  $n$  is colored with a coloring from the first bundle, the second subtree with a coloring from the second bundle, etc. The resulting set of colorings on the depth  $n + 1$  tree, has the following properties: (1) all colorings have the same colors at the leafs, and (2) the number (or fraction) of colorings in this set with a specific color at the root can be computed exactly using only the types of the bundles in the  $d$ -ple. The resulting set of colorings on the depth  $n + 1$  tree can be split again into bundles of the allowed types. The specific recipe for splitting this set into bundles of course will influence to what extent the balanced and near-balanced bundles dominate the entire collection of bundles.

In the special case that the bundles are defined to be the set of colorings with a given boundary, i.e. all types of bundles are allowed and no splitting of bundles occurs, this gives exact recursive computation of the distribution of the marginal distribution at the root.

In this next section we formalize and generalize the above construction.

#### 3.1 Definitions

Let  $\mathcal{V}$  denote a real vector space of finite dimension. Let  $\mathcal{S} = (S_1, \dots, S_n)$  be a finite sequence with  $S_i \in \mathcal{V}$ . We denote the convex hull of  $\mathcal{S}$  in  $\mathcal{V}$  by  $\langle \mathcal{S} \rangle$ . Let  $\alpha_1, \dots, \alpha_n$  be functions from the convex hull  $\langle \mathcal{S} \rangle$  to  $[0, 1]$  such that the following properties hold for every  $\eta \in \langle \mathcal{S} \rangle$ :



1.  $\sum_{i=1}^n \alpha_i(\eta) = 1$ ,
2.  $\eta = \sum_{i=1}^n \alpha_i(\eta) S_i$ .

Thus for every  $\eta \in \langle S \rangle$  these functions define a convex decomposition of  $\eta$ . We will call the tuple  $(S, \alpha_1, \dots, \alpha_n)$  a *skeleton* in  $\mathcal{V}$ , and  $S$  the *base set* of the skeleton.

By  $\mathbf{P}$  we denote a distribution on  $\mathcal{V}$  with finite support as well as a random vector chosen from this distribution. Let the skeleton  $(S, \alpha_1, \dots, \alpha_n)$  be such that the support of the distribution of  $\mathbf{P}$  lies inside  $\langle S \rangle$ . Let  $\mathbf{C}$  be a random element of  $S$  with the following distribution:  $\mathbb{P}(\mathbf{C} = S_i) = \mathbb{E}[\alpha_i(\mathbf{P})]$ . This is well defined by the first condition on the functions  $\alpha_1, \dots, \alpha_n$ . We will call  $\mathbf{C}$  a *survey* of  $\mathbf{P}$  on the skeleton  $(S, \alpha_1, \dots, \alpha_n)$ .

We say that  $\mathbf{A}$  is a survey of  $\mathbf{B}$  without specifying the skeleton, whenever there exists a skeleton with respect to which  $\mathbf{A}$  is the survey of  $\mathbf{B}$ .

### 3.2 Properties of surveys

In this section we show several useful properties of surveys.

**Theorem 5** *If  $\mathbf{P}$  is a distribution on  $\mathcal{V}$  with finite support,  $\mathbf{C}$  is a survey of  $\mathbf{P}$ , and  $\mathbf{D}$  is a survey of  $\mathbf{C}$ , then  $\mathbf{D}$  is a survey of  $\mathbf{P}$ .*

**Proof:** Suppose  $\mathbf{C}$  is a survey of  $\mathbf{P}$  on a skeleton  $(S, \alpha_1, \dots, \alpha_n)$ , and  $\mathbf{D}$  is a survey of  $\mathbf{C}$  on a skeleton  $(T, \beta_1, \dots, \beta_m)$ . Let  $\gamma_i(\eta) := \sum_{j=1}^n \alpha_j(\eta) \beta_i(S_j)$  for  $i = 1, \dots, m$ . Then  $(T, \gamma_1, \dots, \gamma_m)$  is a valid skeleton, because for every  $\eta \in \mathcal{V}$

$$\begin{aligned} \sum_{i=1}^m \gamma_i(\eta) &= \sum_{i=1}^m \sum_{j=1}^n \alpha_j(\eta) \beta_i(S_j) = \sum_{j=1}^n \alpha_j(\eta) \sum_{i=1}^m \beta_i(S_j) = \sum_{j=1}^n \alpha_j(\eta) = 1, \\ \sum_{i=1}^m \gamma_i(\eta) T_i &= \sum_{i=1}^m \sum_{j=1}^n \alpha_j(\eta) \beta_i(S_j) T_i = \sum_{j=1}^n \alpha_j(\eta) \sum_{i=1}^m \beta_i(S_j) T_i = \sum_{j=1}^n \alpha_j(\eta) S_j = \eta. \end{aligned}$$

Finally, we can verify that  $\mathbf{D}$  is a survey of  $\mathbf{P}$  on the skeleton  $(T, \gamma_1, \dots, \gamma_m)$ .

$$\begin{aligned} \mathbb{P}(\mathbf{D} = T_i) &= \mathbb{E}[\beta_i(\mathbf{C})] = \sum_{j=1}^n \mathbb{P}(\mathbf{C} = S_j) \beta_i(S_j) = \sum_{j=1}^n \mathbb{E}[\alpha_j(\mathbf{P})] \beta_i(S_j) \\ &= \mathbb{E} \left[ \sum_{j=1}^n \alpha_j(\mathbf{P}) \beta_i(S_j) \right] = \mathbb{E}[\gamma_i(\mathbf{P})] \end{aligned}$$

■

**Theorem 6** *Let  $\mathbf{P}_1$  and  $\mathbf{P}_2$  be independent distributions on  $\mathcal{V}$  with finite support, and let  $\mathbf{C}_1$  and  $\mathbf{C}_2$  be their surveys. Suppose a distribution  $\mathbf{P}$  on  $\mathcal{V}$  is defined to be equal to  $\mathbf{P}_1$  with probability  $p \geq 0$  and  $\mathbf{P}_2$  with probability  $1 - p$ , and similarly  $\mathbf{C}$  is defined to be  $\mathbf{C}_1$  with probability  $p$  and  $\mathbf{C}_2$  with probability  $1 - p$ . Then  $\mathbf{C}$  is a survey of  $\mathbf{P}$ .*

**Proof:** Suppose  $\mathbf{C}_1$  and  $\mathbf{C}_2$  are surveys respectively of  $\mathbf{P}_1$  and  $\mathbf{P}_2$  on skeletons  $(\mathcal{S}, \alpha_1, \dots, \alpha_n)$ , and  $(\mathcal{T}, \beta_1, \dots, \beta_m)$ . Let  $\mathcal{R} = \mathcal{S} \cap \mathcal{T}$  and  $|\mathcal{R}| = k$ . Without loss of generality, let's suppose that  $\mathcal{R} = \{S_1, \dots, S_k\}$  and  $S_i = T_i$  for  $i = 1, \dots, r$ . Let's denote the union of the two basis sets as  $\mathcal{U} = (U_1, \dots, U_{m+n-k}) = (S_1, \dots, S_n, T_{r+1}, \dots, T_m)$ . This set will be the basis of the skeleton. Next we define the functions:

$$\gamma_j(\eta) := \frac{1}{\mathbb{P}(\mathbf{P} = \eta)} \times \begin{cases} p \alpha_j(\eta) \mathbb{P}(\mathbf{P}_1 = \eta) + (1-p) \beta_j(\eta) \mathbb{P}(\mathbf{P}_2 = \eta) & j = 1, \dots, r \\ p \alpha_j(\eta) \mathbb{P}(\mathbf{P}_1 = \eta) & j = r+1, \dots, n \\ (1-p) \beta_{j-(n-r)}(\eta) \mathbb{P}(\mathbf{P}_2 = \eta) & j = n+1, \dots, m+n-k \end{cases}$$

It is straightforward to check that  $(\mathcal{U}, \gamma_1, \dots, \gamma_{m+n-k})$  is a valid skeleton. To check that  $\mathbf{C}$  is a survey of  $\mathbf{P}$  on this skeleton we just need  $\mathbb{P}(\mathbf{C} = U_j) = \mathbb{E}[\gamma_j(\mathbf{P})]$ . For  $j = 1, \dots, r$

$$\begin{aligned} \mathbb{P}(\mathbf{C} = U_j) &= p \mathbb{P}(\mathbf{C}_1 = U_j) + (1-p) \mathbb{P}(\mathbf{C}_2 = U_j) \\ &= p \mathbb{E}[\alpha_j(\mathbf{P}_1)] + (1-p) \mathbb{E}[\beta_j(\mathbf{P}_2)] \\ &= \sum_{\eta \in \mathcal{P}} (p \mathbb{P}(\mathbf{P}_1 = \eta) \alpha_j(\eta) + (1-p) \mathbb{P}(\mathbf{P}_2 = \eta) \beta_j(\eta)) \\ &= \sum_{\eta \in \mathcal{P}} \mathbb{P}(\mathbf{P} = \eta) \gamma_j(\eta) \\ &= \mathbb{E}[\gamma_j(\mathbf{P})]. \end{aligned}$$

Similarly, for  $j = r+1, \dots, n$ , and for  $j = n+1, \dots, n+m-k$  we have respectively

$$\begin{aligned} \mathbb{P}(\mathbf{C} = U_j) &= p \mathbb{P}(\mathbf{C}_1 = U_j) = p \mathbb{E}[\alpha_j(\mathbf{P}_1)] = \mathbb{E}[\gamma_j(\mathbf{P})], \\ \mathbb{P}(\mathbf{C} = U_j) &= (1-p) \mathbb{P}(\mathbf{C}_2 = U_{j-(n-r)}) = (1-p) \mathbb{E}[\beta_{j-(n-r)}(\mathbf{P}_2)] = \mathbb{E}[\gamma_j(\mathbf{P})], \end{aligned}$$

■

The next theorem is the one that allows us to use surveys in the context of the reconstruction recursion of Theorem 4. Let  $\mathcal{V}_1, \dots, \mathcal{V}_r$  denote real vector spaces of finite dimensions. We say that a function  $f : \mathcal{V}_1 \times \dots \times \mathcal{V}_r \rightarrow \mathcal{V}$  is *multi-affine* if for every  $\eta_1 \in \mathcal{V}_1, \eta_2 \in \mathcal{V}_2, \dots, \eta_r \in \mathcal{V}_r$ ,  $a \geq 0$  and  $b \geq 0$ , such that  $a + b = 1$ ,  $i \in \{1, \dots, r\}$ , and  $\eta'_i \in \mathcal{V}_i$ , it holds that

$$f(\eta_1, \dots, a\eta_i + b\eta'_i, \dots, \eta_r) = af(\eta_1, \dots, \eta_i, \dots, \eta_r) + bf(\eta_1, \dots, \eta'_i, \dots, \eta_r).$$

Recall also that for a vector  $\eta \in \mathcal{V}$  we denote by  $\|\eta\|$  the sum of the coordinates of  $\eta$ .

**Theorem 7** *Let  $f : \mathcal{V}_1 \times \dots \times \mathcal{V}_r \rightarrow \mathcal{V}$  be a multi-affine function. Let  $\mathbf{P}_1, \dots, \mathbf{P}_r$  be independent distributions on  $\mathcal{V}_1, \dots, \mathcal{V}_r$  with finite support, and  $\mathbf{C}_1, \dots, \mathbf{C}_r$  be their respective surveys. If  $\mathbf{Q}$  and  $\mathbf{D}$  are random elements of  $\mathcal{V}$  defined in the following way:*

$$\mathbb{P}(\mathbf{Q} = \eta) \propto \mathbb{E}[\|f(\mathbf{P}_1, \dots, \mathbf{P}_r)\| \times \text{Ind}[f(\mathbf{P}_1, \dots, \mathbf{P}_r) \propto \eta]],$$

$$\mathbb{P}(\mathbf{D} = \eta) \propto \mathbb{E}[\|f(\mathbf{C}_1, \dots, \mathbf{C}_r)\| \times \text{Ind}[f(\mathbf{C}_1, \dots, \mathbf{C}_r) \propto \eta]],$$

*then  $\mathbf{D}$  is a survey of  $\mathbf{Q}$ .*

**Proof:** We first demonstrate the proof for the case  $r = 1$ . We denote  $\mathbf{P}_1$  by  $\mathbf{P}$  and  $\mathbf{C}_1$  by  $\mathbf{C}$ .

Suppose  $\mathbf{C}$  is a survey of  $\mathbf{P}$  on the skeleton  $(\mathcal{S}, \alpha_1, \dots, \alpha_n)$ . Consider the set  $\{f(S_i)/\|f(S_i)\| : S_i \in \mathcal{S}\}$ . The size of this set can be less than  $n$  if for two different indexes  $1 \leq i < i' \leq n$  it happens that  $f(S_i)/\|f(S_i)\| = f(S_{i'})/\|f(S_{i'})\|$ . Let's denote by  $\mathcal{T} = (T_1, \dots, T_m)$  the ordered list of distinct elements corresponding to the above set. Let  $I_j = \{i : f(S_i)/\|f(S_i)\| = T_j\}$ .

Since the support of  $\mathbf{C}$  is  $\mathcal{S}$  it follows that the support of  $\mathbf{D}$  is contained in  $\mathcal{T}$ . It suffices to find  $\beta_1, \dots, \beta_m$  non-negative functions on  $\mathcal{V}$  such that  $\mathbb{P}(\mathbf{D} = T_j) = \mathbb{E}[\beta_j(\mathbf{Q})]$ . We begin with the left-hand side:

$$\begin{aligned} \mathbb{P}(\mathbf{D} = T_j) &\propto \sum_{i=1}^n \mathbb{P}(\mathbf{C} = S_i) \times \|f(S_i)\| \times \mathbb{I}nd[f(S_i)/\|f(S_i)\| = T_j] \\ &= \sum_{i \in I_j} \mathbb{P}(\mathbf{C} = S_i) \times \|f(S_i)\| \\ &= \sum_{i \in I_j} \mathbb{E}[\alpha_i(\mathbf{P})] \times \|f(S_i)\| \\ &= \mathbb{E} \left[ \sum_{i \in I_j} \alpha_i(\mathbf{P}) \|f(S_i)\| \right] \end{aligned}$$

Thus the normalization constant for the above probability is:

$$\sum_{j=1}^m \mathbb{E} \left[ \sum_{i \in I_j} \alpha_i(\mathbf{P}) \|f(S_i)\| \right] = \mathbb{E} \left[ \sum_{i=1}^n \alpha_i(\mathbf{P}) \|f(S_i)\| \right] = \mathbb{E} \left[ \left\| f \left( \sum_{i=1}^n \alpha_i(\mathbf{P}) S_i \right) \right\| \right] = \mathbb{E}[\|f(\mathbf{P})\|],$$

where the second equality follows from the fact that  $f$  is multi-affine.

Next, we look at the right hand-side of the desired equality. For every  $\eta \in \mathcal{V}$  we define

$$\begin{aligned} W(\eta) &= \mathbb{E}[\|f(\mathbf{P})\| \times \mathbb{I}nd[f(\mathbf{P}) \propto \eta]], \\ W &= \mathbb{E}[\|f(\mathbf{P})\|]. \end{aligned}$$

Then by the definition of  $\mathbf{Q}$ ,  $\mathbb{P}(\mathbf{Q} = \eta) = W(\eta)/W$ . For every  $j \in \{1, \dots, m\}$ , let's define

$$\beta_j(\eta) = \frac{1}{W(\eta)} \mathbb{E} \left[ \sum_{i \in I_j} \alpha_i(\mathbf{P}) \times \|f(S_i)\| \times \mathbb{I}nd[f(\mathbf{P}) \propto \eta] \right].$$

It is easy to check that  $\sum_{j=1}^m \beta_j(\eta) = 1$ , and  $\sum_{j=1}^m \beta_j(\eta) T_j = \eta$  for every  $\eta \in \mathcal{V}$ . Finally,

$$\begin{aligned} \mathbb{E}[\beta_j(\mathbf{Q})] &= \sum_{\eta \in \text{supp}(\mathbf{Q})} \mathbb{P}(\mathbf{Q} = \eta) \times \frac{1}{W(\eta)} \mathbb{E} \left[ \sum_{i \in I_j} \alpha_i(\mathbf{P}) \times \|f(S_i)\| \times \text{Ind}[f(\mathbf{P}) \propto \eta] \right] \\ &= \sum_{\eta \in \text{supp}(\mathbf{Q})} \frac{W(\eta)}{W} \times \frac{1}{W(\eta)} \mathbb{E} \left[ \sum_{i \in I_j} \alpha_i(\mathbf{P}) \times \|f(S_i)\| \times \text{Ind}[f(\mathbf{P}) \propto \eta] \right] \\ &= \frac{1}{W} \mathbb{E} \left[ \sum_{i \in I_j} \alpha_i(\mathbf{P}) \|f(S_i)\| \right] = \mathbb{P}(\mathbf{D} = T_j) \end{aligned}$$

Next we generalize the proof to  $r > 1$ . Suppose  $\mathbf{C}_1, \dots, \mathbf{C}_r$  are surveys on skeletons  $(\mathcal{S}^1, \alpha_1^1, \dots, \alpha_{n_1}^1), \dots, (\mathcal{S}^r, \alpha_1^r, \dots, \alpha_{n_r}^r)$ . The same proof applies by defining the base set for the skeleton to be

$$\{f(S_{i_1}^1, \dots, S_{i_r}^r) / \|f(S_{i_1}^1, \dots, S_{i_r}^r)\| : S_{i_1}^1 \in \mathcal{S}^1, \dots, S_{i_r}^r \in \mathcal{S}^r\}$$

and changing the notation to  $I_j = \{(i_1, \dots, i_r) : f(S_{i_1}^1, \dots, S_{i_r}^r) / \|f(S_{i_1}^1, \dots, S_{i_r}^r)\| = T_j\}$ , and  $\mathbf{P} = (\mathbf{P}_1, \dots, \mathbf{P}_r)$ . We have that

$$\mathbb{P}[\mathbf{D} = T_j] = \frac{\mathbb{E} \left[ \sum_{(i_1, \dots, i_r) \in I_j} \left( \prod_{k=1}^r \alpha_{i_k}^k(\mathbf{P}_k) \right) \|f(S_{i_1}^1, \dots, S_{i_r}^r)\| \right]}{\mathbb{E}[\|f(\mathbf{P})\|]} = \mathbb{E}[\beta_j(\mathbf{Q})],$$

where

$$\beta_j(\eta) = \frac{1}{W(\eta)} \mathbb{E} \left[ \sum_{(i_1, \dots, i_r) \in I_j} \left( \prod_{k=1}^r \alpha_{i_k}^k(\mathbf{P}_k) \right) \times \|f(S_{i_1}^1, \dots, S_{i_r}^r)\| \times \text{Ind}[f(\mathbf{P}) \propto \eta] \right].$$

■

**Theorem 8** *Let  $g$  be a convex function on  $\mathcal{V}$ . For every distribution  $\mathbf{P}$  on  $\mathcal{V}$  of finite support, and  $\mathbf{C}$  a survey of  $\mathbf{P}$ ,  $\mathbb{E}[g(\mathbf{P})] \leq \mathbb{E}[g(\mathbf{C})]$  and the equality holds if  $g$  is an affine function.*

**Proof:**

$$\begin{aligned} \mathbb{E}[g(\mathbf{C})] &= \sum_{i=1}^n \mathbb{P}[\mathbf{C} = S_i] g(S_i) = \sum_{i=1}^n \mathbb{E}[\alpha_i(\mathbf{P})] g(S_i) \\ &= \mathbb{E} \left[ \sum_{i=1}^n \alpha_i(\mathbf{P}) g(S_i) \right] \geq \mathbb{E} \left[ g \left( \sum_{i=1}^n \alpha_i(\mathbf{P}) S_i \right) \right] = \mathbb{E}[g(\mathbf{P})] \end{aligned}$$

■

### 3.3 Using surveys to bound the probability of reconstruction

Using the theorems from the previous section now we can show how to calculate recursively a survey of the residual distribution at the root of a random MRF on a skeleton of bounded size. Assume first that the degree distributions and potential function distributions for all levels have small support. Suppose we have a survey  $\mathbf{C}$  for the random MRF of depth  $n$ . For level  $n + 1$  first we calculate, for each possible instantiation of the degree  $r$  and potential functions  $\Psi_1, \dots, \Psi_r$  at the root, a survey of the residual distribution at the root using  $r$  copies of  $\mathbf{C}$  and the recursion of Theorem 4. The result is a survey of the true residual distribution for this instantiation by Theorem 7. Suppose the probability of this instantiation is  $p(r, \Psi_1, \dots, \Psi_r)$ . Next, we combine these distributions by defining a distribution equal to the surveys of each of the instantiations with probability  $p(r, \Psi_1, \dots, \Psi_r)$ . This is a survey of the true residual distribution for the random MRF of depth  $n + 1$  by Theorem 6. Finally, if the support of the resulting survey is bigger than the required bound, we can choose a smaller skeleton and compute a survey of the survey, which by Theorem 5 is also a survey of the true residual distribution.

Finally, since distance from a fixed distribution is a convex function, by Theorem 8 the expected distance of the survey of the residual distribution from  $\pi$  is an upper bound on the distance of the true residual distribution from  $\pi$ . The quality of this bound of course depends on how the skeletons were chosen at each step.

If the bound on the skeleton size is  $b$ , the maximum degree possible for the tree is  $\Delta$ , and the support of the potential-function distribution for every level is at most  $k$ , then the complexity of the computation of the survey is  $O(n(kb)^\Delta)$ . The exponential complexity in  $\Delta$  is in practice prohibitive, because in order to obtain surveys that give good bounds for the probability of reconstruction,  $b$  may have to be large. However, the important improvement here is that while the exact computation is exponential in  $n$ , computing the surveys takes time linear in  $n$ .

A few more remarks regarding implementation are in order:

1. It is not impossible to handle the case when the degree distribution has infinite support. The cases of the small degrees can be computed as above, and for large degrees, if the residual distribution is known to be symmetric (for example if the potential functions are symmetric), then one can use the trivial survey, the one whose basis set is the set of basis vectors. The uniform distribution on the basis vectors is a survey of every symmetric distribution.
2. The size of the domains of the variables also influences the complexity of the algorithm. The computation of the function  $f$  in the recursive step in general requires time  $|D_{v_1}| \times \dots \times |D_{v_r}|$ . However the more important factor is that the size of the skeleton may have to grow significantly with the size of the domain in order to obtain the desired bound. This dependence will have to be studied in the context of particular models.
3. The strategy of choosing the skeleton at every step crucially influences the quality of the bounds. It may be that one type of skeleton is beneficial in the beginning iterations and a different type in the later ones. In our application we used small base sets in the beginning, which makes the first iterations faster, and refined the base sets (increased their size) progressively. Perhaps strategies for choosing the skeleton can be designed based on the current distribution (such as sampling a few probability vectors from it), but we have not found such a general-purpose strategy that performs well.

4. In order to obtain rigorous results, the computation of the survey of the residual distribution has to be carried out with rational numbers. However, naturally the size of the denominators increases exponentially, thus a “rounding” step is required at every iteration. In the case of symmetric distributions, such as those generated in the Potts model, this can be handled in a similar way to item 1. We choose a large  $N$  which will be a bound on the size of the denominators. At every step the weights of all the vectors in the survey are rounded down to the nearest allowed rational number, one with denominator  $N$ , and the remaining weight is distributed among the basis vectors. The resulting distribution is a survey of the original one. By the symmetry, in the resulting distribution all the basis vectors have the same weight, therefore their denominator is at most  $|D_v|N$ .

## 4 Application to the Potts model on small-degree trees

Let  $T$  be a random infinite tree rooted at the vertex  $\rho$  such that the number of children  $d$  of every vertex is distributed according to a random variable  $\mathbf{d}$  with expected value  $\bar{d}$  and maximum possible value  $d_{max}$ . Let the domain of values that can be assigned to each vertex be denoted by  $D = \{1, \dots, q\}$ , and we will also call these values *colors*. The channel  $M$  on each edge in the Potts model is given by

$$M_{i,j} = \begin{cases} 1 - p & \text{if } i = j, \\ \frac{p}{q-1} & \text{otherwise,} \end{cases}$$

where  $0 \leq p \leq 1$ . This channel corresponds to the  $q$ -state Potts model on the tree. Denote the resulting configurations of the tree by  $\bar{\sigma}$  and the alphabet at a vertex  $v$  by  $\sigma_v$ . The Potts model weighs the resulting configurations according to the Hamiltonian function  $H(\bar{\sigma}) = \sum_{(u,v) \in E(T)} \mathbb{I}nd[\sigma_u = \sigma_v]$  which counts the number of edges in which the color on both end points is the same. On a finite tree, the probability distribution is given by

$$\mathbb{P}(\bar{\sigma}) = \frac{1}{Z} \exp \left( \beta \sum_{(u,v) \in E(T)} \mathbb{I}nd[\sigma_u = \sigma_v] \right)$$

where  $Z$  is a normalizing constant and  $\beta$  is an inverse temperature parameter of the Potts model.

The second largest eigenvalue of the matrix  $M$  is denoted by

$$\lambda = 1 - \frac{pq}{q-1} = \frac{e^\beta - 1}{e^\beta + q - 1}.$$

In line with the terminology for the Potts model,  $\lambda < 0$  corresponds to the *ferromagnetic regime* while  $\lambda > 0$  corresponds to the *anti-ferromagnetic regime*. The special case of proper colorings corresponds to  $\lambda = \frac{-1}{q-1}$ .

The *branching number* of an infinite tree is the supremum of the real numbers  $\gamma \geq 1$  such that  $T$  admits a positive flow from the root to infinity, where on every edge  $e$ , the flow is bounded by  $\gamma^{-\ell_e}$ , where  $\ell_e$  denotes the number of edges, including  $e$  on the path from  $e$  to the root. Note that for the  $d$ -ary tree, the branching number is  $d$ . In the case where each vertex of the tree has  $k$  children with probability  $p_k$ , it is known that if the expected number of children  $m = \sum_k k p_k > 1$ , then the branching number is  $m$  almost surely (see [7]).

The Kesten-Stigum bound [6] for the reconstruction problem says that for a tree with branching number  $d$  such that  $\lambda d^2 > 1$  reconstruction holds. For the Potts model, Mossel and Peres [12] have shown that non-reconstruction holds if

$$d \frac{q\lambda^2}{2 + \lambda(q-2)} < 1$$

and this bound was improved in [8].

At  $q = 3$ , for large enough degree, the Kesten-Stigum bound was recently shown to be sharp in both the ferromagnetic and antiferromagnetic cases [16]. However, there is still a gap for small degrees. We consider the anti-ferromagnetic and the ferromagnetic Potts models  $q = 3$ , on 2- and 3-ary trees, and on the tree in which every vertex is chosen with equal probability to be 2 or 3, and show bounds on the threshold for non-reconstruction.

Let  $\sigma$  denote a random configuration of a tree  $T$  given by the transition matrix. Recall that for  $a \in D$ ,  $L^a(n)$  denotes the random coloring of  $L(n)$  conditioned on  $\sigma_\rho = a$ . In agreement with the notation of [16] we define the random variable

$$X^+(n) = \mathbb{P}_{L \sim L^a(n)}(\sigma_\rho = a | L).$$

This is the conditional probabilities of the color  $a$  at the root when the coloring of the vertices at level  $n$  is chosen conditioned on  $\sigma_\rho = a$ . Note that by the symmetry of the channel, the distribution of  $X^+(n)$  does not depend on the particular  $a \in D$ .

Let  $Y^+(n) := X^+(n) - \frac{1}{q}$  and denote

$$x_n = \mathbb{E}[Y^+(n)], \quad \text{and} \quad z_n = \mathbb{E}[Y^+(n)^2].$$

Here, the expectation is taken over the randomness of the tree and the Markov process on the tree (the random coloring). We go back to the unconditional distribution using the following identity of Sly [16]:

**Claim 9 ([16])** *The following relations hold:*

$$x_n = \mathbb{E} \left[ \sum_{a=1}^q \left( \mathbb{P}(\sigma_\rho = a | L(n)) - \frac{1}{q} \right)^2 \right] \geq z_n \geq 0$$

Therefore, clearly, the condition

$$\lim_{n \rightarrow \infty} x_n = 0$$

is equivalent to non-reconstruction and further, if for each  $a$ ,

$$\mathbb{E} \left[ \left( \mathbb{P}(\sigma_\rho = a | L(n)) - \frac{1}{q} \right)^2 \right] \leq \varepsilon,$$

then  $x_n \leq q\varepsilon$ .



By expanding the recursion for the expectation  $\mathbb{E}[Y^+(n)]$  using a Taylor expansion, exactly as in [16], we obtain a bound on  $x_{n+1}$  in terms of  $x_n$  and  $z_n$ . Extending the analysis of [16] to random trees is straightforward. First, we obtain a relation in terms of a fixed degree  $d$  and then take the expectation over the distribution for the degree  $\mathbf{d}$ .

**Theorem 10** *Let  $q, \mathbf{d}, \bar{d}, d_{max}, \lambda, x_n$  and  $z_n$  be defined as above. Then,*

$$x_{n+1} \leq \bar{d}\lambda^2 x_n + \sum_{j=2}^{d_{max}} E\left[\binom{\mathbf{d}}{j}\right] \lambda^{2j} \left[ \frac{2(q-1)}{q^2} \left( \frac{q(q-3+\lambda)}{q-1} x_n - \frac{q^2\lambda}{q-1} z_n \right)^j \right. \quad (6)$$

$$+ \frac{(q-1)(q-2)}{q^2} \left( -\frac{q(3q-6+2\lambda)}{(q-1)(q-2)} x_n + \frac{2q^2\lambda}{(q-1)(q-2)} z_n \right)^j \quad (7)$$

$$\left. - \frac{2(q-1)}{q} \left( -\frac{q}{q-1} x_n \right)^j \right]. \quad (8)$$

**Proof:** The degree of the root of the random tree of depth  $n+1$  is  $\mathbf{d}$ . We first bound the expectation conditional on  $\mathbf{d} = d$ . Following the calculations of [16], we obtain

$$\begin{aligned} \mathbb{E}[Y^+(n+1)|\mathbf{d} = d] &\leq \frac{2(q-1)}{q^2} \left( 1 + \frac{\lambda^2 q(q-3+\lambda)}{q-1} x_n - \frac{q^2\lambda^3}{q-1} z_n \right)^d \\ &\quad + \frac{(q-1)(q-2)}{q^2} \left( 1 - \frac{\lambda^2 q(3q-6+2\lambda)}{(q-1)(q-2)} x_n + \frac{2q^2\lambda^3}{(q-1)(q-2)} z_n \right)^d \\ &\quad - \frac{2(q-1)}{q} \left( 1 - \frac{\lambda^2 q}{q-1} x_n \right)^d + 1 - 1/q \\ &= d\lambda^2 x_n + \sum_{j=2}^d \binom{d}{j} \lambda^{2j} \left[ \frac{2(q-1)}{q^2} \left( \frac{q(q-3+\lambda)}{q-1} x_n - \frac{q^2\lambda}{q-1} z_n \right)^j \right. \\ &\quad + \frac{(q-1)(q-2)}{q^2} \left( -\frac{q(3q-6+2\lambda)}{(q-1)(q-2)} x_n + \frac{2q^2\lambda}{(q-1)(q-2)} z_n \right)^j \\ &\quad \left. - \frac{2(q-1)}{q} \left( -\frac{q}{q-1} x_n \right)^j \right]. \end{aligned}$$

Taking the expectation over the degree we obtain the statement of the theorem. ■

We obtain the following Corollary by applying Theorem 10 to particular values of  $q$  and degree distributions  $\mathbf{d}$ . The inequalities are obtained by optimizing each term in the summation (6) separately subject to the constraint that  $0 \leq z_n \leq x_n$ .

**Corollary 11** *Let  $q = 3$ . In the ferromagnetic regime,*

1. If  $\mathbf{d} = 2$  with probability 1, and  $\lambda > 0$

$$x_{n+1} \leq 2\lambda^2 x_n + \frac{3}{2}\lambda^4 x_n^2 (2\lambda^2 + 4\lambda + 1)$$

2. If  $\mathbf{d} = 3$  with probability 1, and  $\lambda > 0$

$$x_{n+1} \leq 3\lambda^2 x_n + \frac{9}{2}\lambda^4 x_n^2 (2\lambda^2 + 4\lambda + 1) - \frac{9}{2}\lambda^6 x_n^3 (1 + \lambda)^3$$

3. If  $\mathbf{d} = 2$  with probability 1/2,  $\mathbf{d} = 3$  with probability 1/2, and  $\lambda > 0$

$$x_{n+1} \leq \frac{5}{2}\lambda^2 x_n + 3\lambda^4 x_n^2 (2\lambda^2 + 4\lambda + 1) - \frac{9}{4}\lambda^6 x_n^3 (1 + \lambda)^3$$

In the antiferromagnetic regime,

4. If  $\mathbf{d} = 3$  with probability 1, and  $\lambda = -\frac{1}{2}$

$$x_{n+1} \leq \frac{3}{4}x_n + \frac{63}{32}x_n^2 - \frac{351}{256}x_n^3$$

Using the above bounds, we show that non-reconstruction holds in the following cases.

**Theorem 12** Let  $q = 3$ . In the ferromagnetic regime,

1. if  $\mathbf{d} = 2$  with probability 1, non-reconstruction holds for  $\lambda \leq 0.69$ ;
2. if  $\mathbf{d} = 3$  with probability 1, reconstruction holds for  $\lambda \leq 0.555$ ;
3. if  $\mathbf{d} = 2$  with probability 1/2,  $\mathbf{d} = 3$  with probability 1/2, non-reconstruction holds for  $\lambda \leq 0.61$ .

In the antiferromagnetic regime,

4. If  $\mathbf{d} = 3$  with probability 1, and  $\lambda = -\frac{1}{2}$  (the case of proper colorings), there is non-reconstruction.

**Proof:** Let  $x^* = x^*(\lambda, q, d)$  denote the upper bound on  $x_n$  given by the algorithm when it is run with inputs  $\lambda, q, d$ . The values obtained for  $x^*$  by the algorithm were as follows:

- If  $\mathbf{d} = 2$  w.p. 1,  $\lambda = 0.69$ ,  $x^* = 0.02939\dots$
- If  $\mathbf{d} = 3$  w.p. 1,  $\lambda = 0.555$ ,  $x^* = 0.04457\dots$
- If  $\mathbf{d} = 2$  or 3 each w.p. 1/2,  $\lambda = 0.61$ ,  $x^* = 0.04057\dots$
- If  $\mathbf{d} = 3$  w. p. 1 and  $\lambda = -1/2$ ,  $x^* = 0.00038\dots$

$d$	KS [6]	MP [12]	FK [4]	Theorem 12
2	0.7071..	0.6666..	0.7018..	0.69
3	0.5773..	0.5302..	0.5731..	0.555
2.5	0.6324..	0.5873..	0.6278..	0.61

Table 1: The Kesten-Stigum upper bound on the non-reconstruction threshold and the values of  $\lambda$  up to which non-reconstruction has been shown in [12], [4] and here, for  $q = 3$  in the ferromagnetic regime ( $\lambda > 0$ ).

The values  $x^*$  are an upper bound on  $x_n$ . It can be verified by substituting the  $\lambda$  above into the corresponding inequalities in Corollary 11 that in each case  $x_{n+1} < Cx_n$  where  $C$  is a constant smaller than 1. This implies non-reconstruction since in the limit  $x_n$  goes to 0. ■

For each of these cases the algorithm was run using Maple 12 and with only integer computations, using the rounding procedure described in the previous section. The base sets of the skeletons were selected manually, refining them whenever the resulting bounds stop improving. The decomposition functions of the skeletons were selected to minimize the expected total variation distance between the true vector and its decomposition using the LP solver of Maple 12. Not more than 100 iterations were needed for every case to obtain the required bound. The implementation was run on a MacBook with a 2GHz Intel Core 2 Duo processor and 1GB of RAM. The limiting factor is that the last tens of iterations take hours because the skeleton size we choose towards the end is close to 100 (in the case of  $d=2$ , 200). It is reasonable to expect that with more computational power or time each of these bounds can be improved, although going beyond the bounds of Formentin and Külske, if they are not tight, may require significantly better resources.

The values of  $\lambda$  we obtain for non-reconstruction in the ferromagnetic regime are shown in Table 4 for a comparison with previous bounds from [6, 12, 4] (the bounds of [8] are not explicitly derived, so we have not included them). The second column is the Kesten-Stigum bound below which reconstruction is known to hold. In all cases we improve the bound of [12]. The anti-ferromagnetic case (3-coloring on the 3-ary tree) is also not implied by the bounds of [12].

## Acknowledgements

The authors would like to thank Allan Sly and Florent Krzakala for discussions on the topic. E.M. also thanks the Kavli Institute of Theoretical Physics China for a productive stay during the initial stages of this work.

## References

- [1] N. BHATNAGAR, J. VERA, E. VIGODA, AND D. WEITZ, *Reconstruction for colourings on trees*, preprint, Available at <http://arxiv.org/abs/0711.3664>, (2008).

- [2] C. BORGS, J. CHAYES, E. MOSSEL, AND S. ROCH, *The Kesten-Stigum reconstruction bound is tight for roughly symmetric binary channels*, in Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2006, pp. 518–530.
- [3] W. EVANS, C. KENYON, Y. PERES, AND L. J. SCHULMAN, *Broadcasting on trees and the Ising model*, Ann. Appl. Probab., 10 (2000), pp. 410–433.
- [4] M. FORMENTIN AND C. KÜLSKE, *On the Purity of the free boundary condition Potts measure on random trees*. ArXiv:0810.0677, 2008.
- [5] H. GEORGII, *Gibbs Measures and Phase Transitions*, Vol 9, Studies in Mathematics, de Gruyter, 1988.
- [6] H. KESTEN AND B. P. STIGUM, *Additional limit theorems for indecomposable multidimensional Galton-Watson processes*, Ann. Math. Statist., 37 (1966), pp. 1463–1481.
- [7] R. LYONS, *Random walks and percolation on trees*, Ann. Probab., 18 (1990), pp. 931–958.
- [8] F. MARTINELLI, A. SINCLAIR, AND D. WEITZ, *Fast mixing for independent sets, colorings, and other models on trees*, Random Structures and Algorithms, 31 (2007), pp. 134–172. Submitted to Communication in Mathematical Physics. Extended abstract appeared in proceedings of 44'th FOCS.
- [9] M. MÉZARD AND A. MONTANARI, *Reconstruction on trees and the spin glass transition*, J. Stat. Phys., 124 (2006), pp. 1317–1350.
- [10] M. MÉZARD, G. PARISI, AND R. ZECCHINA, *Analytic and algorithmic solution of random satisfiability problems*, Science, 297 (2002), pp. 812–815. (Scienceexpress published on-line 27-June-2002; 10.1126/science.1073287).
- [11] E. MOSSEL, *Reconstruction on trees: beating the second eigenvalue*, Ann. Appl. Probab., 11 (2001), pp. 285–300.
- [12] E. MOSSEL AND Y. PERES, *Information flow on trees*, Ann. Appl. Probab., 13 (2003), pp. 817–844.
- [13] T. RICHARDSON AND R. URBANKE, *The capacity of low-density parity check codes under message-passing decoding*, IEEE Trans. Info. Theory, 47 (2001), pp. 599–618.
- [14] ———, *Modern coding theory*, Cambridge University Press, 2008.
- [15] A. SLY, *Reconstruction of random colourings*, Communications of Mathematical Physics, (2008).
- [16] ———, *Reconstruction of symmetric Potts Models*, in Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC), 2009.
- [17] L. ZDEBOROVA AND F. KRZAKALA, *Phase transitions in the coloring of random graphs*, Phys. Rev. E., 76 (2007), p. 031131.